Robust Multiview Subspace Learning With Nonindependently and Nonidentically Distributed Complex Noise

Zongsheng Yue, Hongwei Yong, Deyu Meng, *Member, IEEE*, Qian Zhao, Yee Leung, and Lei Zhang, *Senior Member, IEEE*

***Abstract*— Multiview Subspace Learning (MSL), which aims at obtaining a low-dimensional latent subspace from multiview data, has been widely used in practical applications. Most recent MSL approaches, however, only assume a simple independent identically distributed (i.i.d.) Gaussian or Laplacian noise for all views of data, which largely underestimates the noise complexity in practical multiview data. Actually, in real cases, noises among different views generally have three specific characteristics. First, in each view, the data noise always has a *complex* configuration beyond a simple Gaussian or Laplacian distribution. Second, the noise distributions of different views of data are generally *nonidentical* and with evident distinctiveness. Third, noises among all views are *nonindependent* but obviously correlated. Based on such understandings, we elaborately construct a new MSL model by more faithfully and comprehensively considering all these noise characteristics. First, the noise in each view is modeled as a Dirichlet process (DP) Gaussian mixture model (DPGMM), which can fit a wider range of complex noise types than conventional Gaussian or Laplacian. Second, the DPGMM parameters in each view are different from one another, which encodes the "nonidentical" noise property. Third, the DPGMMs on all views share the same high-level priors by using the technique of hierarchical DP, which encodes the "nonindependent" noise property. All the aforementioned ideas are incorporated into an integrated graphics model which can be appropriately solved by the variational Bayes algorithm. The superiority of the proposed method is verified by experiments on 3-D reconstruction simulations, multiview face modeling, and background subtraction, as compared with the current state-of-the-art MSL methods.**

***Index Terms*— Dirichlet process (DP) mixture model, hierarchical Dirichlet process (HDP), multiview, subspace learning, variational Bayes.**

Manuscript received August 7, 2018; revised January 29, 2019; accepted April 29, 2019. Date of publication June 19, 2019; date of current version April 3, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1004300 and in part by China NSFC under Project 61661166011, Project 11690011, Project 61603292, Project 61721002, and Project U1811461. *(Zongsheng Yue and Hongwei Yong contributed equally to this work.) (Corresponding author: Deyu Meng.)*

Z. Yue, D. Meng, and Q. Zhao are with the Institute for Information and System Sciences, Ministry of Education Key Lab of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: dymeng@mail.xjtu.edu.cn).

H. Yong and L. Zhang are with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong.

Y. Leung is with the Department of Geography and Resource Management, Institute of Future Cities, The Chinese University of Hong Kong, Hong Kong.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2019.2917328

I. Introduction

CONSIDERABLE data used in real applications, such as face recognition [1], video surveillance [2], background subtraction [3], and 3-D structure reconstruction [4]–[6], are collected from various domains or extracted with diverse features in order to capture possibly comprehensive underlying information of data. Such kinds of data are known as multiview data. A typical example is shown in Fig. 1(a), where a single scene is captured by multiple cameras located at different positions.

Since the information obtained only from one view of data is relatively insufficient to describe the full shape of an object [7], it is always advantageous to leverage the information of multiple views to obtain a better description. Multiview learning has thus been attracting much research attention recently in both academic and practical fields [8], and various methods have been proposed in recent years [9]. Multiview subspace learning (MSL) [7], [10]–[12] represents one of the most typical categories of approaches along this research line.

The basic assumption of MSL is that each view is generated from one latent subspace, which is shared by all views. Most current MSL methods aim to embed the data into this low-dimensional latent subspace with a set of dictionaries such that each view of the data could be linearly combined through the coefficients and corresponding dictionary [7], [10], [11], [13], [14]. Through such a learning mechanism, many of the current methods can achieve good performance on multiview data collected from ideal scenarios.

However, one apparent drawback of these current methods is that they generally specify a simple L_2-norm or L_1-norm loss in the model, implying that the noise across all views of data is assumed to be an independent identically distributed (i.i.d.) Gaussian or Laplacian distribution. This, however, always deviates from the real noise configurations in practical cases. Specifically, noises in real multiview data always possess the following threefold characteristics.

First, the noise in each view is always too ***complex*** to be simply modeled as a Gaussian or Laplacian distribution. Like the surveillance multiview videos shown in Fig. 1(d), the noise (i.e., residual besides the stable background) in each view can be decomposed into multiple modalities (i.e., the foreground objects, the shadow of the objects, and some camera noises),

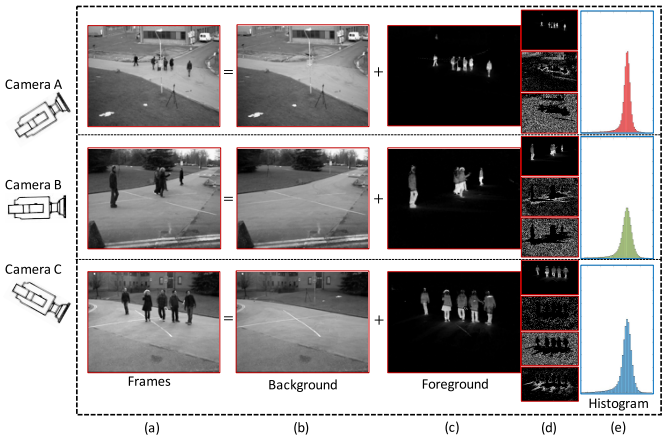


Fig. 1. (a) Representative frames of original video captured from the similar scene in three different views. (b)–(e) Backgrounds, foregrounds (residuals), Gaussian components of the residuals, and residual histograms of all views extracted by the proposed NIID-MSL method at the same scale for all panels.

which need to be finely depicted by more complex noise models beyond Gaussian or Laplacian [15]–[17].

Second, in practical cases, noises in different data views are ***nonidentically*** distributed, i.e., they are always of evident distinctiveness due to different collecting angles, domains or sources. For example, as shown in Fig 1(e), the residuals' histograms of different views are obviously different, implying that noises of different views should have different distributions. The negligence to such inter-view noise difference tends to impact adversely on the performance of current methods.

Third, for real multiview data, the noise distributions are always ***nonindependent*** and correlations among different views are evident [18]. For instance, for multiple surveillance videos captured by cameras with different angles of the same scene, the foreground objects incline to appear or leave simultaneously across all views of the videos, making the noise types among all views evidently correlated, as can be observed from the common noise shapes among views shown in Fig. 1(c) and (e). By taking such noise correlations among all data views into consideration should more faithfully reflect the noise characteristics of multiview data and enhance the robustness of an MSL method for practical applications under complicated noises.

To address the aforementioned noise fitting issues, in this paper, we propose a new MSL method that fully takes the noise characteristics into consideration. The main contribution of this paper can be summarized as follows.

1) Instead of assuming simple Gaussian or Laplacian noise-like conventional studies, we model the noise in each view of data as a Dirichlet process (DP) Gaussian mixture model (DPGMM), which is capable of fitting wider range of noise types [3], [19], [20] beyond the Gaussian or Laplacian distribution, thus finely encoding the "complex" noise property underlying data.
2) Compared with the traditional i.i.d. noise assumption, we initiate a hierarchical model for modeling the non-i.i.d. noise in multiview data. The DPGMM noise parameters in each view are different from others in order to encode the "nonidentical" noise property. Furthermore, the DPGMM parameters among all views are shared from the similar high-level DP parameters via the technique of Hierarchical DP (HDP), which encodes the "nonindependent" noise property. Under this modeling strategy, the noise in practical multiview data is more faithfully represented so as to enhance the robustness of the MSL method in dealing with real complex noises.
3) The proposed model corresponds to a nonparametric Bayesian generative model, which is capable of automatically adapting the noise complexity based on data. A variational Bayes algorithm is readily designed to solve the model, and each of the involved parameters can be effectively updated in closed-form. Compared with the state-of-the-art MSL methods, experimental results substantiate the superiority of our proposed method.

This paper is organized as follows. Section II introduces related works on MSL. Section III provides some preliminaries on the HDP technique. Section IV presents the main model against the investigated problem, and the variational Bayes algorithm is given in Section V. Experimental results on synthetic and real data sets are demonstrated in Section VI, and this paper is rounded up with a conclusion in Section VII.

II. Related Work

During the past decades, many MSL approaches have been proposed. The canonical correlation analysis (CCA) [21], [22], which exploits the shared latent subspace across diverse views, is a typical fundamental work along this line of research. Subsequently, a series of related approaches based on CCA have been proposed. For instance, Bach and Jordan [23] provided a probabilistic interpretation of CCA and perfected the theoretical basis of CCA. To handle nonlinear alignment, the kernel CCA [24] was proposed by projecting data onto a high-dimensional feature space. In addition, sparse formulation of CCA was also proposed in [25]. To enable the method to perform in complex noises, Nicolaou *et al.* [26] adopted L_1 loss to enhance the robustness of CCA, and Bach and Jordan [23] put forward a Student-t density model to handle more outliers.

Several other methods have also been presented to deal with the MSL task. The shared Gaussian latent variable model (sGP LVM) [27], [28] learns the common latent structure of multiview data by the Gaussian process. Jia *et al.* [10] imposed structured sparsity to MSL through solving two convex subproblems alternately. Similarly, White *et al.* [11] proposed a more general unified framework under arbitrary convex loss function for MSL and then reformulated the problem to obtain a more precise solution of the problem. Other related methods on convex formulations for MSL can be found in [14] and [29]–[31]. Cauchy loss [32] was first put forward as a more robust estimator for MSL in [7]. By simultaneously considering the correlation and independence of multiview data, some approaches divided the data into correlated components among all views and specific components with respect to each view. For example, JIVE [33] decomposed multiview data into the sum of three terms: a low-rank approximation capturing the joint structure among different views, low-rank approximations capturing individual structures for each view

and residual noise. Motivated by JIVE, Zhou *et al.* [34] used a common orthogonal basis extraction (COBE) algorithm to identify and separate the shared and individual features.

However, most traditional MSL methods generally assume a simple form for the noise in multiview data, like i.i.d. Gaussian or Laplacian distribution, which always largely deviates from the noise configuration in practical multiview data and, thus, tends to have an unstable performance in the presence of non-i.i.d. complex noises. Thus, our aim is to more faithfully capture and represent real noise structures to alleviate this robustness issue.

III. Preliminaries

In this section, we review some preliminary knowledge about DP and HDP to set the stage for the subsequent presentation of our model.

A. Dirichlet Process

The DP, introduced by Ferguson [35], is a distribution over distributions. The DP is parameterized by a base distribution and a concentration parameter. Specifically, let H be a probability distribution over any measurable set Θ and γ be a positive real number. Then, we say G is a DP distributed with the base distribution H and concentration parameter γ, written as $G \sim \mathrm{DP}(\gamma, H)$, if

$$(G(A_1), \cdots, G(A_r)) \sim \mathrm{Dir}(\alpha H(A_1), \cdots, \alpha H(A_r)) \quad (1)$$

for every finite measurable partitions $A_1, A_2, \ldots, A_r$ of Θ.

There is another more explicit characterization of DP in terms of a stick-breaking construction due to [36]. Considering two infinite collections of independent random variables, $\pi_i' \sim \mathrm{Beta}(1, \gamma)$ and $\eta_i^* \sim H$, for $i = 1, 2, \ldots$, then the stick-breaking representation of G is as follows:

$$\pi_i = \pi_i' \prod_{j=1}^{i-1}(1 - \pi_j'), \quad G = \sum_{i=1}^{\infty} \pi_i \delta_{\eta_i^*} \quad (2)$$

where $\delta_{\eta_i^*}$ represents the Dirac delta function concentrated at η_i^*. Sethuraman [36] showed that G as defined in this way is a random probability measure according to $\mathrm{DP}(\gamma, H)$.

The DP Gaussian mixture clustering model is a typical example of the DP models. It is an advanced version of conventional GMM, with the capability of adaptively adjusting its Gaussian component number based on the data. The main idea of the model is as follows: setting H as a Gaussian-inverse-Wishart distribution [conjugate priors of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$], thus G is a distribution over an infinite number of clusters, each corresponding to a pair value $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. A draw from G [$G \sim \mathrm{DP}(\gamma, H)$] will choose one cluster k and return one pair value $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, which determines one specific Gaussian component. Thus, for every data sample needed to be clustered, we can sample from the posterior of G until convergence is reached. We can also understand DP in the sense of the stick-breaking construction. Instead of sampling the Gaussian component (i.e., cluster) directly for each data sample, it completes the clustering process through sampling the component weight π_k of the Gaussian mixture for each cluster. In this clustering process, the value of the concentration parameter γ controls the final number of clusters. Through this stochastic dynamical mechanism, DP offers full flexibility in selecting the number of clusters (i.e., the number of Gaussian components).

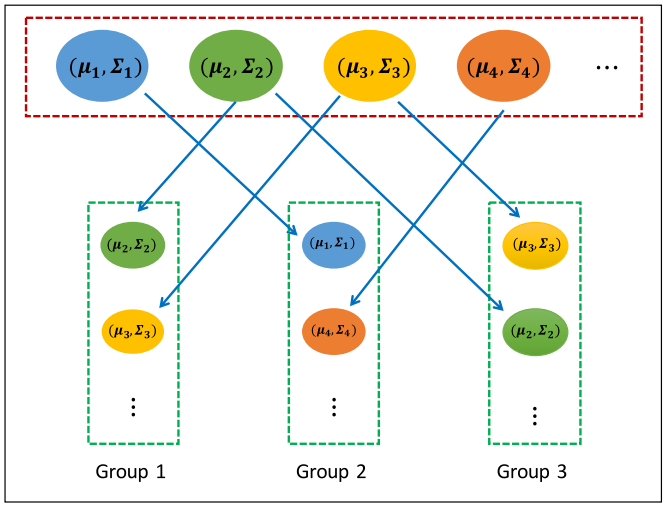


Fig. 2. Illustration of the hierarchical clustering process through HDP. The ellipses with different colors denote different clusters with parameter pairs $(\boldsymbol{u}_k, \boldsymbol{\Sigma}_k)$. The green dotted rectangles, representing the clustering result of each group data, share cluster parameters from a top-level clustering result in the red dotted rectangle. The number of clusters at the two-levels is both dynamically adjusted during the clustering process until it converges.

B. Hierarchical Dirichlet Process

The HDP [37] is originally proposed as a hierarchical Bayesian model for natural language processing, which provides a flexible framework for sharing mixture components among groups of related data. A two-level HDP is a collection of DPs that share a base distribution H, and meanwhile, these DPs are also drawn from a higher level DP, that is,

$$G \sim \mathrm{DP}(\gamma, H), \quad G_j \sim \mathrm{DP}(\alpha_j, G) \quad (3)$$

where j is an index for groups of data. It should be noted that all distributions G_j possess different parameters while share the base distribution G, which can thus be regarded as a desirable noise prior readily representing our expected "nonindependently and nonidentically distributed" noise distribution. We can extend the DPGMM example in Section III-A to a clustering problem with three groups of related data, and then the clusters in each group share the parameters from a similar high-level clustering process, which leads to a hierarchical clustering process as shown in Fig. 2.

IV. MSL Model With Non-i.i.d. Complex Noise

A. Notations

For the convenience of formulating our model, we first introduce some necessary notations in the following. We use light lowercase letters, bold lowercase letters, and bold uppercase letters to denote scalars, vectors, and matrices, respectively. Given a matrix $\boldsymbol{X}$, we use the term $\boldsymbol{x}_{\cdot j}$ to denote its jth column vector, $\boldsymbol{x}_{i\cdot}$ its ith row vector, and x_{ij} the element in its ith row and jth column, respectively. For probability distributions,

$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\mathcal{N}(\mu, \xi^{-1})$ denotes the univariate Gaussian distribution with mean μ and precision ξ, Beta(a, b) denotes the beta distribution with parameters a and b, Gam(c, d) denotes the Gamma distribution with shape parameter c and scale parameter d, and Multi($\boldsymbol{\pi}$) denotes the multinomial distribution with parameters $\boldsymbol{\pi}$.

B. Model Formulation

Now, we give the formulation of our proposed model. Let the observed multiview data be $\boldsymbol{\mathcal{X}} = \{\boldsymbol{X}^v\}_{v=1}^V$, where $\boldsymbol{X}^v \in \mathbb{R}^{d\times n}$ represents the vth view of data, and n and d are the number of data samples and the dimensionality in each view, respectively.[1] By considering a generative model, we can decompose the observed data into

$$\boldsymbol{X}^v = \boldsymbol{S}^v + \boldsymbol{E}^v \tag{4}$$

where $\boldsymbol{E}^v = \{e_{ij}^v\}_{d\times n}$ denotes the residual term (i.e., noise component) and $\boldsymbol{S}^v \in \mathbb{R}^{d\times n}$ is the expected data located on the latent subspace.

1) Non-i.i.d. Noise Modeling: Instead of using simple i.i.d. unimodal distribution (Gaussian or Laplacian) to model the residual noise in all views of data, we model the noise of each as a DPGMM, which is an advanced version of nonparametric GMM model capable of adaptively rectifying the number of Gaussian components based on data [38], [39], and then associate these DPGMMs by considering the intrinsic correlations among all views. This can be achieved by a two-level HDP as described in Section III-B.

Specifically, we consider the noise of each view $\boldsymbol{E}^v$. Following the idea of [19] and [20], we can model it as an elementwise Gaussian mixture distribution

$$e_{ij}^v \sim \sum\nolimits_{k=1}^{K} \pi_k^v \mathcal{N}\left(0, {\xi_k}^{-1}\right) \tag{5}$$

which can be equivalently reformulated as a two-level generative process with a latent variable $\boldsymbol{b}^v \in \mathbb{R}^{d\times n}$

$$e_{ij}^v \sim \mathcal{N}\left(0, \xi_{b_{ij}^v}^{-1}\right), \quad b_{ij}^v \sim \text{Multi}(\boldsymbol{\pi}^v) \tag{6}$$

where $\boldsymbol{\pi}^v \in \mathbb{R}^K$, $b_{ij}^v \in \{1, 2, \cdots, K\}$. In order to encode the correlation and distinctiveness of noise in different views, we adopt a hierarchical Dirichlet distribution prior, that is,

$$\begin{aligned} \boldsymbol{\pi}^v &\sim \text{Dir}(\alpha^v \boldsymbol{\beta}), \\ \boldsymbol{\beta} &\sim \text{Dir}(\gamma/K, \gamma/K, \cdots, \gamma/K). \end{aligned} \tag{7}$$

To make the model flexible, we further assume $K \to +\infty$, Teh *et al.* [37] showed that the limit of (5)–(7) is the following HDP noise model:

$$\begin{aligned} G &\sim DP(\gamma, H), \quad G^v \sim \text{DP}(\alpha^v, G) \\ \psi_{ij}^v &\sim G^v \quad e_{ij}^v \sim \mathcal{N}(0, {\psi_{ij}^v}^{-1}) \end{aligned} \tag{8}$$

where $H = \text{Gam}(e_0, f_0)$.

[1]For notation convenience, we assume each view is with the same dimensionality. However, our method can also be easily used in cases where dimensionalities are different in multiple views.

Alternatively, we can intuitively explain our noise model (8) from another hierarchical Bayesian model equivalently based on the stick-breaking construction. Instead of placing a prior on $\boldsymbol{\beta}$, each view can choose a subset of T mixture components from a high-level set of K mixture components as shown in Fig. 2.

(1) Draw K (K is sufficiently large) samples from the Gamma distribution, that is,

$$\xi_k \sim \text{Gam}(e_0, f_0)$$

and form K Gaussian distributions $\mathcal{N}(0, \xi_k^{-1})$ as listed in the red dotted rectangle in Fig. 2.

(2) T ($T \le K$) values are selected from $\{\xi_k\}_{k=1}^K$ for each view, whose index $\boldsymbol{c}^v \in \mathbb{R}^T$ is constructed by the following stick-breaking representation

$$c_t^v \sim \text{Multi}(\boldsymbol{\beta}), \ \ \beta_k = \beta_k' \prod\nolimits_{l=1}^{k-1}(1-\beta_l'), \ \ \beta_k' \sim \text{Beta}(1, \gamma)$$

where $c_t^v \in \{1, 2, \cdots, K\}$, $\boldsymbol{\beta} \in \mathbb{R}^K$. Thus, we can get T Gaussian distributions $\mathcal{N}(0, \xi_{c_t^v}^{-1})$ as listed in the green dotted rectangles in Fig. 2.

(3) Generate the indicated variable z_{ij}^v from another stick-breaking process

$$\begin{aligned} z_{ij}^v \sim \text{Multi}(\boldsymbol{\pi}^v), \quad \pi_t^v = \pi_t^{v\prime} \prod_{l=1}^{t-1}(1-\pi_l^{v\prime}), \\ \pi_t^{v\prime} \sim \text{Beta}(1, \alpha^v) \end{aligned}$$

where $z_{ij}^v \in \{1, 2, \cdots, T\}$, $\boldsymbol{\pi}^v \in \mathbb{R}^T$.

(4) Generate noise for each view of data, that is,

$$e_{ij}^v \sim \mathcal{N}\left(0, \xi_{c_{z_{ij}^v}^v}^{-1}\right).$$

Combining the above steps (1)–(4), we can get the following hierarchical noise model represented by two coupled stick-breaking process

$$\begin{aligned} &\xi_k \sim \text{Gam}(e_0, f_0), \quad e_{ij}^v \sim \mathcal{N}\left(0, \left(\xi_{c_{z_{ij}^v}^v}\right)^{-1}\right) \\ &c_t^v \sim \text{Multi}(\boldsymbol{\beta}), \quad z_{ij}^v \sim \text{Multi}(\boldsymbol{\pi}^v) \\ &\beta_k = \beta_k' \prod_{l=1}^{k-1}(1-\beta_l'), \quad pi_t^v = \pi_t^{v\prime} \prod_{l=1}^{t-1}(1-\pi_l^{v\prime}) \\ &\beta_k' \sim \text{Beta}(1, \gamma), \quad \pi_t^{v\prime} \sim \text{Beta}(1, \alpha^v). \end{aligned} \tag{9}$$

It should be noted that the $\boldsymbol{\xi}$ in (9) and the $\boldsymbol{\psi}$ in (8) have the following relation:

$$\xi_{c_{z_{ij}^v}^v} = \psi_{ij}^v.$$

The concentration parameters, $\boldsymbol{\alpha}$ and γ, mainly affect the number of Gaussian components of the second-level GMM in each view and the first-level GMM for the entire data set, respectively. Based on the above stick-breaking representation, we place the conjugate Gamma distribution on them [40], that is,

$$\gamma \sim \text{Gam}(m_0, n_0), \quad \alpha^v \sim \text{Gam}(g_0, h_0). \tag{10}$$

In the above formulations, e_0, f_0, g_0, h_0, m_0, and n_0 are the hyperparameters of the prior distributions, which are all easily set as small values to make the prior noninformative.

Note that such noise modeling [i.e., (8) or (9)] implies that each view of data has its own DPGMM noise, which is not identical, while these DPs come from the same higher lever DP, and share the same hyperparameters, which are thus not independent. Therefore, this noise model nicely delivers the non-i.i.d. noise property underlying the multiview data.

2) Latent Subspace Modeling: As conventional MSL methods [7], [10], we assume that each view of data is embedded into a latent subspace $\boldsymbol{R}$ with a dictionary $\boldsymbol{L}^v$. Based on such an assumption, we formulate $\boldsymbol{S}^v \in \mathbb{R}^{d\times n}$ as the product of two smaller matrices $\boldsymbol{L}^v \in \mathbb{R}^{d\times l}$ and $\boldsymbol{R} \in \mathbb{R}^{l\times n}$, that is,

$$\boldsymbol{S}^v = \boldsymbol{L}^v \boldsymbol{R} = \sum_{r=1}^{l} \boldsymbol{L}^v_{\cdot r} \boldsymbol{R}_{r\cdot} \tag{11}$$

In order to obtain a full Bayesian model, we impose the following priors on $\boldsymbol{R}$ [20]

$$\boldsymbol{R}_{r\cdot} \sim \mathcal{N}\left(\boldsymbol{0}, \frac{1}{\tau_r} I_n\right). \tag{12}$$

In addition, we also impose column priors on $\boldsymbol{L}^v$ [10], [41], that is,

$$\boldsymbol{L}^v_{\cdot r} \sim \mathcal{N}\left(\boldsymbol{0}, \frac{1}{\lambda^v_r} I_d\right). \tag{13}$$

The intuition behind this formulation is that we excepted each $\boldsymbol{S}^v$ to only depend on a subset of latent dimensions, and the conjugate priors imposed on the precision variables λ^v_r and τ_r are

$$\lambda^v_r \sim \text{Gam}(a_0, b_0), \quad \tau_r \sim \text{Gam}(c_0, d_0) \tag{14}$$

respectively, where a_0, b_0, c_0, and d_0 are all hyperparameters.

3) Full Bayesian Model of NIID-MSL: Combining (4) and (9)–(14) together, we can construct a full Bayesian model of MSL with non-i.i.d. noise. We, thus, call our model NIID-MSL in brief. The corresponding graphical model of our method is shown in Fig. 3. The goal turns to infer the posteriors of all involved variables

$$p(\boldsymbol{L}, \boldsymbol{R}, \boldsymbol{\xi}, \boldsymbol{C}, \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \gamma \,|\boldsymbol{\mathcal{X}}) \tag{15}$$

where $\boldsymbol{C} = \{c^v_t\}$, $\boldsymbol{Z} = \{z^v_{ij}\}$.

C. Discussion

Our proposed NIID-MSL model considers the correlation and distinctiveness of the noises among different views, which is inspired by the general nonidentical and nonindependent noise characteristics of practical multiview data. It should be noted that the proposed method can be easily transformed into two of its degenerated versions, including: 1) i.i.d. noise version, by using a single DPGMM to fit the noise of all views and 2) the nonidentical but independent noise version by assuming a different DPGMM for each view. Due to the page limitation, we put the models of both such degenerated versions in the supplementary material.[2] In addition, we also list experimental results in this supplementary file to show the necessity of considering such non-i.i.d. assumptions in our noise model.

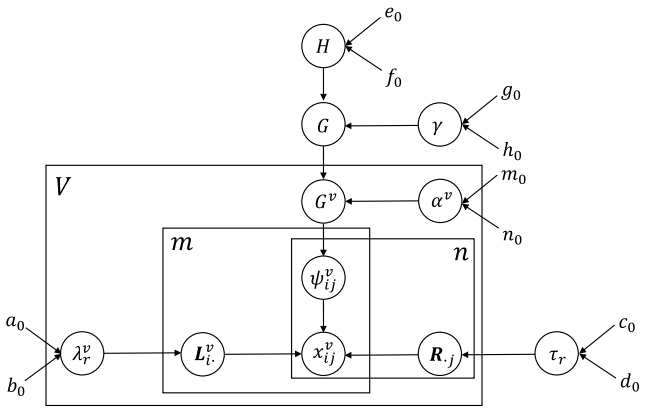


Fig. 3. Graphical model of NIID-MSL.

V. Variational Inference

Directly computing the posterior distribution under a HDP mixture prior is intractable, thus approximate inference methods are required to be designed. Although Markov chain Monte Carlo (MCMC) sampling method can provide a very accurate approximation to the posterior [42], this method is limited for massive-scale data by its high computational cost and the difficulty to detect convergence. As a more efficient and deterministic alternative to MCMC, we adopt a variational Bayesian (VB) method for NIID-MSL in this paper [43].

VB seeks an approximation distribution $q(\boldsymbol{x})$ to the true posteriors $p(\boldsymbol{x}|\mathcal{D})$ ($\mathcal{D}$ is the observed data) by solving the following variational optimization problem

$$\min_{q\in\mathcal{C}} KL(q||p) = -\int q(\boldsymbol{x}) \ln\left\{\frac{p(\boldsymbol{x}|\mathcal{D})}{q(\boldsymbol{x})}\right\} d\boldsymbol{x} \tag{16}$$

where $KL(q||p)$ denotes the KL divergence between $q(\boldsymbol{x})$ and $p(\boldsymbol{x}|\mathcal{D})$, and $\mathcal{C}$ denotes the set of probability densities with certain restrictions to make the minimization tractable. Assuming $q(\boldsymbol{x}) = \prod_i q_i(x_i)$, the closed-form solution to $q_j(x_j)$ with other factors fixed, can be attained as

$$q^*_j(x_j) = \frac{\exp\{E_{\boldsymbol{x}\setminus x_j}[\ln p(\boldsymbol{x}, \mathcal{D})]\}}{\int \exp\{E_{\boldsymbol{x}\setminus x_j}[\ln p(\boldsymbol{x}, \mathcal{D})]\} dx_j} \tag{17}$$

where $E_{\boldsymbol{x}}[\cdot]$ denotes the expectation over the variable $\boldsymbol{x}$, and $\boldsymbol{x}\setminus x_j$ represents the set of $\boldsymbol{x}$ with x_j removed. Equation (16) can be solved by alternatively calculating (17) with respect to all of its involved variables.

A. Estimation of the Posterior

Based on the stick-breaking construction and the truncation strategy [44], we can approximate the posterior

[2]The supplementary material can be downloaded from http://gr.xjtu.edu.cn/c/document_library/get_file?folderId=2589730&name=DLFE-112915.pdf.

distribution (15) with the following factorized form:

$$\begin{aligned} &q(\boldsymbol{L}, \boldsymbol{R}, \boldsymbol{\xi}, \boldsymbol{C}, \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \gamma) \\ &= \prod_{v=1}^{V} \prod_{r=1}^{l} q\left(\boldsymbol{L}_{\cdot r}^{v}\right) q\left(\lambda_{r}^{v}\right) \prod_{r=1}^{l} q(\boldsymbol{R}_{r\cdot}) q(\tau_r) \prod_{k=1}^{K} q(\xi_k) q(\beta_k') \\ &\quad \times \prod_{v=1}^{V} \prod_{i,j}^{d,n} q\left(z_{ij}^{v}\right) \prod_{v=1}^{V} \prod_{t=1}^{T} q v\left(c_{t}^{v}\right) q\left(\pi_{v}^{t\prime}\right) \prod_{v=1}^{V} q(\alpha^{v}) q(\gamma). \end{aligned} \tag{18}$$

Then, we can analytically infer all the factorized distributions involved in (18) as below. The computational details are provided in the supplementary material.

1) Update Noise Component: The parameters involved in the noise components are $\boldsymbol{\xi}$, $\boldsymbol{C}$, $\boldsymbol{Z}$, $\boldsymbol{\pi}$, $\boldsymbol{\beta}$. We use the stick-breaking procedure by setting a large enough value of K and T for truncated approximation, and then get the following updating equations on $\boldsymbol{C}$ and $\boldsymbol{Z}$

$$q(z_{ij}^{v}) = \text{Multi}\left(z_{ij}^{v} | \boldsymbol{\rho}_{ij}^{v}\right), \quad q\left(c_{t}^{v}\right) = \text{Multi}\left(c_{t}^{v} | \boldsymbol{\varphi}_{t}^{v}\right) \tag{19}$$

where

$$\begin{aligned} \rho_{ijt}^{v} &\propto \exp\left(\sum_{k} \varphi_{tk}^{v} E\left[\ln \mathcal{N}\left(e_{ij}^{v} | 0, \xi_k{}^{-1}\right)\right] + E\left[\ln \pi_{t}^{v}\right]\right) \\ \varphi_{tk}^{v} &\propto \exp\left(\sum_{i,j} \rho_{ijt}^{v} E\left[\ln \mathcal{N}\left(e_{ij}^{v} | 0, \xi_k{}^{-1}\right)\right] + E[\ln \beta_k]\right) \\ E\left[\ln \pi_{t}^{v}\right] &= E\left[\ln \pi_{t}^{v\prime}\right] + \sum_{s=1}^{t-1} E\left[\ln(1 - \pi_{s}^{v\prime})\right] \\ E\left[\ln \beta_k\right] &= E[\ln \beta_k'] + \sum_{l=1}^{k-1} E[\ln(1 - \beta_l')]. \end{aligned} \tag{20}$$

As for $\boldsymbol{\xi}$, based on its conjugate priors, we have

$$q(\xi_k) = \text{Gam}(\xi_k | e_k, f_k) \tag{21}$$

where

$$\begin{aligned} e_k &= \frac{1}{2} \sum_{v,i,j,t} \rho_{ijt}^{v} \varphi_{tk}^{v} + e_0, \\ f_k &= \frac{1}{2} \sum_{v,i,j,t} \rho_{ijt}^{v} \varphi_{tk}^{v} E\left[\left(x_{ij}^{v} - \boldsymbol{L}_{i\cdot}^{v\,T} \boldsymbol{R}_{\cdot j}\right)^2\right] + f_0. \end{aligned} \tag{22}$$

Similarly, for $\boldsymbol{\pi}$, $\boldsymbol{\beta}$, we have

$$q\left(\pi_{t}^{v\prime}\right) = \text{Beta}\left(\pi_{t}^{v\prime} | r_{t}^{v}, w_{t}^{v}\right), \; q(\beta_k') = \text{Beta}\left(\beta_k' | s_k^1, s_k^2\right) \tag{23}$$

where

$$\begin{aligned} r_{t}^{v} &= \sum_{i,j} \rho_{ijt}^{v} + 1, \quad w_{t}^{v} = \sum_{i,j,s=t+1} \rho_{ijs}^{v} + E[\alpha^{v}], \\ s_k^1 &= \sum_{v,t} \varphi_{tk}^{v} + 1, \quad s_k^2 = \sum_{v,t,s=k+1} \varphi_{ts}^{v} + E[\gamma]. \end{aligned} \tag{24}$$

2) Update HDP Concentration Parameters: According to the conjugate priors of $\boldsymbol{\alpha}$ and γ in (10), we can get the following updating equations:

$$q(\alpha^{v}) = \text{Gam}(\alpha^{v} | m^{v}, n^{v}), \quad q(\gamma) = \text{Gam}(\gamma | g, h) \tag{25}$$

where

$$\begin{aligned} m^{v} &= T + m_0, \quad n^{v} = n_0 - \sum_{t} E[\ln(1 - \pi_{t}^{v\prime})] \\ g &= K + g_0, \quad h = h_0 - \sum_{k} E[\ln(1 - \beta_k')]. \end{aligned} \tag{26}$$

3) Update the Subspace Components: The parameters involved in the latent subspace component are $\boldsymbol{L}^{v}$, $\boldsymbol{R}$, $\boldsymbol{\lambda}$, $\boldsymbol{\tau}$. For each row of $\boldsymbol{L}^{v}$, using the factorization in (11), we can get

$$q(\boldsymbol{L}_{i\cdot}^{v}) = \mathcal{N}\left(\boldsymbol{L}_{i\cdot}^{v} | \boldsymbol{\mu}_{i}^{v}, \boldsymbol{\Sigma}_{i}^{v}\right) \tag{27}$$

with mean $\boldsymbol{\mu}_{i}^{v}$ and convariance matrix $\boldsymbol{\Sigma}_{i}^{v}$, given by

$$\begin{aligned} \boldsymbol{\Sigma}_{i}^{v} &= \left(\sum_{j,t,k} \rho_{ijt}^{v} \varphi_{tk}^{v} E[\xi_k] E\left[\boldsymbol{R}_{\cdot j} \boldsymbol{R}_{\cdot j}^{T}\right] + \boldsymbol{\Lambda}_{L}^{v}\right)^{-1}, \\ \boldsymbol{\mu}_{i}^{v} &= \boldsymbol{\Sigma}_{i}^{v} \left(\sum_{j,t,k} \rho_{ijt}^{v} \varphi_{tk}^{v} E[\xi_k] x_{ij}^{v} E[\boldsymbol{R}_{\cdot j}]\right) \end{aligned} \tag{28}$$

where $\boldsymbol{\Lambda}_{L}^{v} = \text{diag}(E[\boldsymbol{\lambda}^{v}])$. Similarly, for each column of $\boldsymbol{R}$, we have

$$q(\boldsymbol{R}_{\cdot j}) = \mathcal{N}(\boldsymbol{R}_{\cdot j} | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \tag{29}$$

and the mean $\boldsymbol{\mu}_j$ and the convariance $\boldsymbol{\Sigma}_j$ are calculated by

$$\begin{aligned} \boldsymbol{\Sigma}_j &= \left(\sum_{v,i,t,k} \rho_{ijt}^{v} \varphi_{tk}^{v} E[\xi_k] E\left[\boldsymbol{L}_{i\cdot}^{v} \boldsymbol{L}_{i\cdot}^{v\,T}\right] + \boldsymbol{\Lambda}_{R}\right)^{-1}, \\ \boldsymbol{\mu}_j &= \boldsymbol{\Sigma}_j \left(\sum_{v,i,t,k} \rho_{ijt}^{v} \varphi_{tk}^{v} E[\xi_k] x_{ij}^{v} E[\boldsymbol{L}_{i\cdot}^{v}]\right) \end{aligned} \tag{30}$$

where $\boldsymbol{\Lambda}_{R} = \text{diag}(E[\boldsymbol{\tau}])$.

For parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$, we have

$$q(\lambda_{r}^{v}) = \text{Gam}\left(\lambda_{r}^{v} | a_{r}^{v}, b_{r}^{v}\right), \quad q(\tau_r) = \text{Gam}\left(\tau_r | c_r, d_r\right) \tag{31}$$

with

$$\begin{aligned} a_{r}^{v} &= \frac{d}{2} + a_0, \quad b_{r}^{v} = \frac{1}{2} E\left[\boldsymbol{L}_{\cdot r}^{v\,T} \boldsymbol{L}_{\cdot r}^{v}\right] + b_0, \\ c_r &= \frac{n}{2} + c_0, \quad d_r = \frac{1}{2} E\left[\boldsymbol{R}_{r\cdot}^{T} \boldsymbol{R}_{r\cdot}\right] + d_0 \end{aligned} \tag{32}$$

where d and n are the dimensionality and the number of the observed data $\boldsymbol{X}^{v}$ in the vth view, respectively.

B. Settings of Hyperparameters

We set all the hyperparameters involved in our model in a noninformative manner to make as small an impact as possible on the inference of the posterior distribution [43]. Specifically, throughout our experiments, we easily set a_0, b_0, c_0, d_0, e_0, f_0, g_0, h_0, m_0, n_0 for a small value 10^{-6}. Our method performs stably well in all experiments with these easy settings.

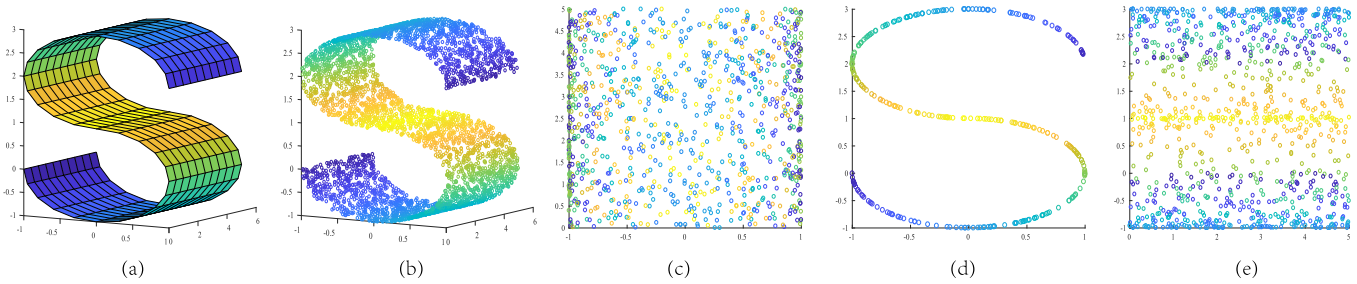


Fig. 4. Illustration of the "S-curve" and its corresponding projection onto the 2-D planes. (a) Original "S-curve." (b) 4000 data points randomly sampled from the "S-curve." (c) Projection onto the xy plane. (d) Projection onto the xz plane. (e) Projection onto the yz plane.

C. VB Algorithm Acceleration

As with the truncated DP [44], there is a tradeoff issue in the designed VB algorithm for the NIID-MSL: the larger K and T are, the more accurate the model becomes, but the slower is the model in inferring the parameters. In order to accelerate the inference speed, we employed two strategies.

1) Cut the components in the first-level GMM: The coefficient β_k in (9) reflects the importance of the kth component, and discard the kth component if $\beta_k < 1e-4$.
2) Merge the mixture components of each view in the second-level GMMs: The value φ_{tk}^{v} in (19) represents the exception that the tth Gaussian component in the vth view is shared from the kth component in the first-level GMM. In each iteration, we merge the t_1-th component and the t_2-th component into one for each view if the following relation is satisfied:

$$\left\|\boldsymbol{\varphi}_{t_1}^{v}-\boldsymbol{\varphi}_{t_2}^{v}\right\|_{\max}<0.01$$

where $\|\cdot\|_{\max}$ means the Max norm of a vector.

VI. Experimental Results

We evaluate the performance of our proposed NIID-MSL model qualitatively and quantitatively on both synthetic and real data sets. Five state-of-the-art MSL methods, including FLSSS [10], MSL [11], CSRL [14], JIVE [33], and MISL [7], are considered for comparison in the synthetic simulation, face image recovery, and background subtraction experiments. All the involved parameters in these comparison methods are finely tuned by default settings or following the rules in their papers to guarantee their possibly optimal performance. The source code of the proposed NIID-MSL is available at https://github.com/zsyOAOA/NIID-MSL.

A. Synthetic Simulations

We evaluate our proposed method on the problem of 3-D structure reconstruction based on a synthetic "S-curve" data set as shown in Fig. 4(a), which is generated in the similar manner as [45]. We uniformly sampled 4000 3-D data points from the "S-curve" manifold [Fig. 4(b)], and then projected them onto three 2-D planes (i.e., xy plane, xz plane, yz plane) as three views of the data, as depicted in Fig. 4(c)–(e). Obviously, each of these three views of data correspond to a 2-D subspace in the 3-D space. Our goal is to reconstruct its 3-D "S-curve" shape by utilizing these three views of the 2-D projected data. When the data are clean and directly generated from the "S-curve" manifold (i.e., Case 1 of Table I), all competing MSL methods can effectively reconstruct the manifold shapes, as displayed in the first row of Fig. 5. To evaluate the robustness of all competing methods, we further added five types of i.i.d. and non-i.i.d. noises with different signal to noise ratios (SNR) to the data to increase its complexity, in which the SNR of the vth view is calculated with the following equation:

$$\mathrm{SNR}^{v}=10\log_{10}\left(\frac{\frac{1}{d\times n}\sum_{i,j}^{d,n}\left(x_{ij}^{v}-\bar{x}\right)^{2}}{{\sigma_{\mathrm{noise}}^{v}}^{2}}\right)$$

where $\bar{x}^{v}=(1/d\times n)\sum_{i,j}^{d,n}x_{ij}^{v}$. Details of the added noises are listed in Table I. It should be noted that Case 2 corresponds to the i.i.d. GMM noise, while the others are non-i.i.d. ones.

TABLE I
SNR Value of the Added Noises in Six Cases of the Synthetic Experiments

Number	view 1	view 2	view 3
Case 1		No noise	
Case 2	20	20	20
Case 3	15	20	20
Case 4	15	20	25
Case 5	10	20	20
Case 6	10	20	30

Two criteria are utilized for performance assessment: 1) relative reconstruction square error (RRSE): $\|\hat{L}-L\|_F/\|L\|_F$, and 2) relative reconstruction absolute error (RRAE): $\|\hat{L}-L\|_1/\|L\|_1$, where L and $\hat{L}$ denote the groundtruth and reconstructed data, respectively. The performance of all competing methods are listed in Table II. We can see from the table that, in the case without noise (Case 1), all competing methods can perform well with small reconstruction error. For all of the other noisy cases, however, our proposed method outperforms the other comparison methods in most cases. Specifically, for the i.i.d. noise case 2, our method performs substantially better than the other competing methods. As the noise becomes more complicated and with evidently more non-i.i.d. configurations, our proposed method performs evidently much better than all others. This superiority substantiates the robustness of our method against complex noise in the MSL task. For easy visualization, Fig. 5 displays the reconstructed 3-D points by

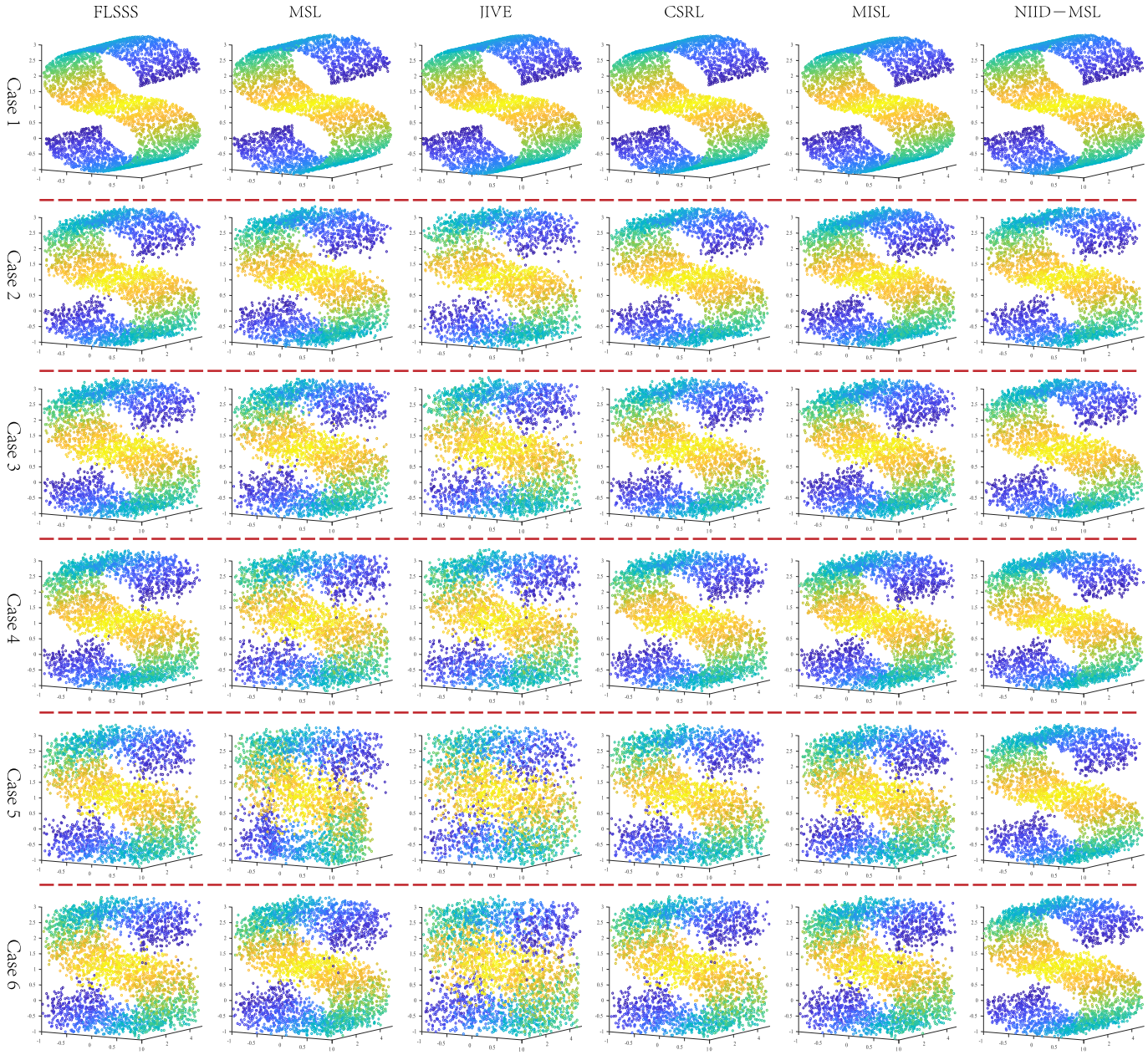


Fig. 5. Illustration of the reconstructed 3-D "S-curve" of different methods. Rows correspond to the reconstructed results of all the methods in cases 1–6 respectively, and each column represents the reconstruction of a specific competing method.

all the competing methods in all cases. It is easy to see that the performances of most competing methods degenerate under complex noise cases, especially in non-i.i.d. cases, while our proposed method can stably attain a more superior reconstruction performance. This complies with the quantitative comparison and further substantiates the robustness of our proposed method.

B. Face Image Recovery Experiments

In this section, we test the effectiveness and the robustness of the proposed NIID-MSL method on multiview face images by using the CMU Multi-PIE face data set [46], which includes 33 7 subjects with multiple poses and expressions. For our purpose, 100 subjects were randomly selected to avoid the "out of memory" issue that tends to occur in the utilized MSL methods, and each subject contains one face image captured at five different angles ($-30°$, $-15°$, $0°$, $15°$, and $30°$) with size 128×96. A typical subject is shown in the first row of Fig. 6(a). The original face images (no noise case) were directly fed into the algorithm to test the effectiveness of our proposed method. Furthermore, we added three types of noises to the original images to test the robustness of different methods as follows:

Gaussian Noise: Zero-mean Gaussian noise was added to each view of images. The noise variances were all set for 0.02 for all views.

Block Noise: Two blocks were randomly selected in every face image for all five views to add "salt and pepper" noise. The density of "salt and pepper" was uniformly sampled from the $[0.5, 0.8]$ region.

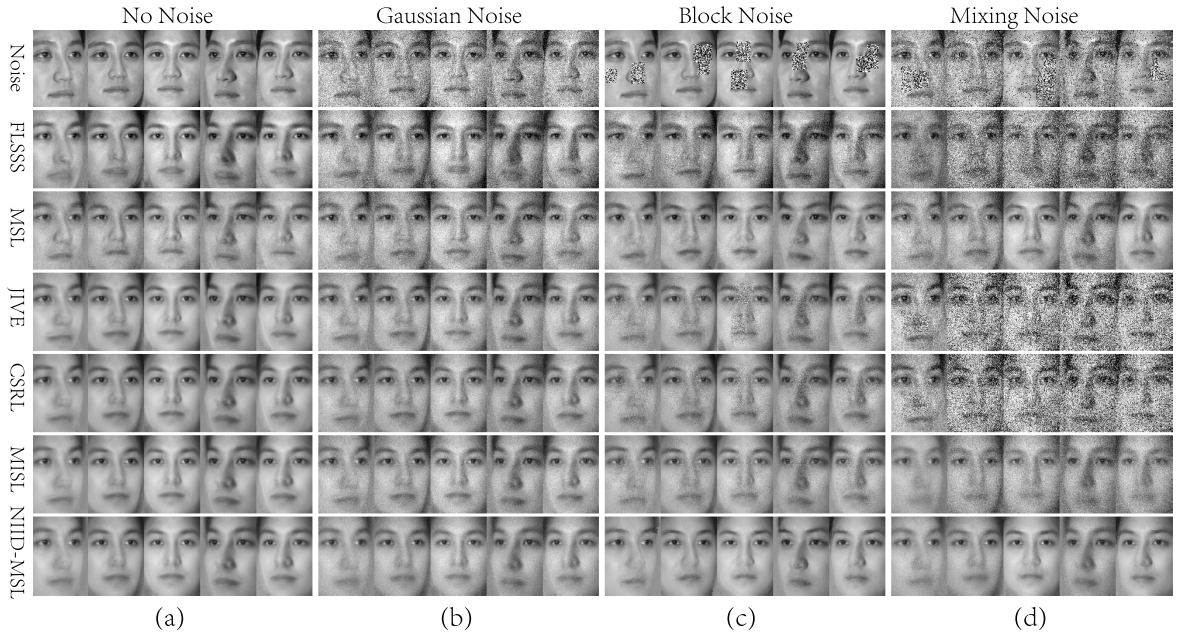


Fig. 6. Typical reconstructed face images of CMU Multi-PIE face data set. Noise faces [or original face in (a)] and reconstructed faces by different methods (top to bottom). (a) Typical reconstructed faces from original face images (left to right). (b)–(d) Typical reconstructed faces under Gaussian noise, block noise, and mixture noise (left to right).

TABLE II

QUANTITATIVE PERFORMANCE COMPARISON OF ALL COMPETING METHODS ON THE "S-CURVE" SYNTHETIC EXPERIMENTS. IN EACH SERIES OF EXPERIMENTS, THE BEST AND THE SECOND BEST RESULTS ARE HIGHLIGHTED IN BOLD AND ITALIC, RESPECTIVELY

Methods		FLSSS	MSL	JIVE	CSRL	MISL	NIID-MSL
Case 1	RRSE	0.016	0.0081	**6.46e-16**	0.010	0.025	*3.20e-9*
	RRAE	0.015	0.011	**5.67e-16**	0.011	0.028	*3.18e-8*
Case 2	RRSE	0.057	0.067	0.084	*0.055*	0.058	**0.054**
	RRAE	0.059	0.068	0.088	*0.057*	0.060	**0.056**
Case 3	RRSE	0.075	0.092	0.11	**0.074**	0.075	**0.074**
	RRAE	0.076	0.092	0.11	**0.075**	0.076	**0.075**
Case 4	RRSE	0.069	0.087	0.093	*0.068*	0.070	**0.054**
	RRAE	0.069	0.082	0.084	*0.068*	0.071	**0.054**
Case 5	RRSE	0.12	0.28	0.16	0.12	*0.12*	**0.065**
	RRAE	0.11	0.22	0.15	0.11	0.11	**0.066**
Case 6	RRSE	0.11	0.12	0.15	0.11	0.11	**0.053**
	RRAE	0.11	0.11	0.11	*0.10*	0.11	**0.052**

Mixing Noise:[3] Two types of noises randomly selected from the Gaussian noise, sparse noise (randomly select 10% of the image pixels and assign random values within the region $[-2,2]$), and block noise, as mentioned above, were added to each view.

For fair comparison, the rank in all experiments is set for 15 for all of the competing methods. The performances are also evaluated in terms of RRSE and RRAE. The results are listed in Table III, and the images recovered by different methods are shown in Fig. 6. From Table III and Fig. 6, we have the following observations.

[3]This "Mixing noise" case simply indicates that we add more than one kind of noise to the original face images, which does not follow the mixture distribution in statistics.

TABLE III

QUANTITATIVE COMPARISON OF ALL COMPETING METHODS ON MULTI-PIE FACE DATA. IN EACH SERIES OF EXPERIMENTS, THE BEST AND THE SECOND BEST RESULTS ARE HIGHLIGHTED IN BOLD AND ITALIC, RESPECTIVELY

Methods		FLSSS	MSL	JIVE	CSRL	MISL	NIID-MSL
No noise	RRSE	0.031	0.015	**0.0074**	0.010	*0.0092*	*0.0092*
	RRAE	0.13	0.090	*0.068*	0.077	0.070	**0.067**
Gaussian noise	RRSE	0.018	0.022	0.019	0.018	*0.017*	**0.014**
	RRAE	0.17	0.12	0.11	*0.10*	*0.10*	**0.09**
Block noise	RRSE	0.019	**0.011**	0.022	0.018	0.019	**0.011**
	RRAE	0.10	**0.068**	0.11	0.10	0.099	*0.070*
Mixture noise	RRSE	0.14	*0.020*	0.17	0.11	0.11	**0.015**
	RRAE	0.23	*0.10*	0.26	0.21	0.16	**0.086**

1) The NIID-MSL method attains the best (six out of eight cases) or second best (two out of eight cases) performance in both quantitative metrics and visualization. Such superiority is especially evident in the more complicated mixture noise case.
2) Most of the state-of-the-art MSL methods can remove the effect of illumination and reconstruct from the original images without noise, which validates the effectiveness of these methods.
3) Under the Gaussian noise case, the L_2 loss methods (JIVE, CSRL, and FLSSS) outperform the L_1 loss method (MSL). In contrast, as for block noise, L_1 loss method establishes a clear superiority. The MISL method based on Cauchy loss also have a good performance, since Cauchy loss has a breakdown point of nearly 50% [32] and thus is more robust. Our proposed NIID-MSL evidently surpasses other methods, which proves its robustness against complex noise.

The promising performance of the NIID-MSL method as shown in these face experiments can be easily explained by

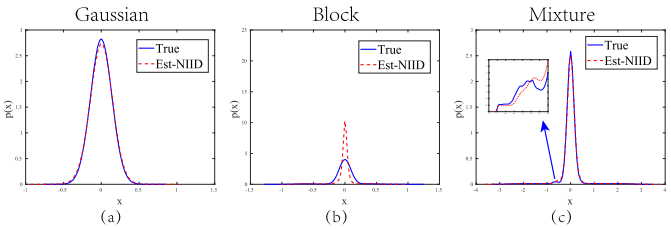


Fig. 7. Visual comparison of the ground truth (denoted by True) noise probability density functions and those estimated (denoted by Est-NIID) by the NIID-MSL method in the face image recovery experiments. Inset depicts the zoomed-in view of the indicated portion. (a) Gaussian noise case. (b) Block noise case. (c) Mixing noise case.

Fig. 7, which compares the ground truth noise distributions and the estimated ones.[4] It can be easily observed that the estimated noise distributions can match the true ones to a very accurate extent for Gaussian and mixture noise. As for block occlusion, reasonably, a peak fat tail distribution, which is similar to the Laplacian distribution, is estimated by our NIID-MSL method, since the noise is only embedded in some local parts of the data. Such good noise-fitting capability of our proposed method then naturally leads to its good performance on its recovery of the ground truth faces from their noisy versions.

C. Foreground Detection on RGB Data

Surveillance videos with RGB color frames are very commonly collected in real life due to the widespread of digital cameras. Intuitively, the information contained in the R, G, and B channels of such kinds of data is highly correlated while also with differences. Here, we attempt to use MSL methods to detect the foreground on the RGB data regarding the three channels of a video as three views of data. We employ the Wallflower data set[5] [47] and I2R data set[6] to evaluate the proposed method. The former includes six video sequences and latter contains eight video sequences, and each video contains hundreds or thousands of frames in total, of which one or several frames are preannotated as the ground truth of the corresponding foreground object masks. Due to memory limitation of our computer, we only crop about 300 frames containing the ones with preannotated masks in each video sequences to test the effectiveness and robustness of different methods. The utilized video sequences are released at https://github.com/zsyOAOA/NIID-MSL for easy reproduction of our experiments.

The residual in such kinds of data is generally non-i.i.d. Specifically, each channel generally has its own different distribution since the sensitive spectral band of R, G, and B sensors is distinct, as shown in Fig. 8. However, all sensors observed the same object, so the residual distributions are also of certain correlation. The noise in RGB surveillance videos is, thus, generally more complicated than those assumed by the current MSL methods, which in better accordance with our non-i.i.d structure.

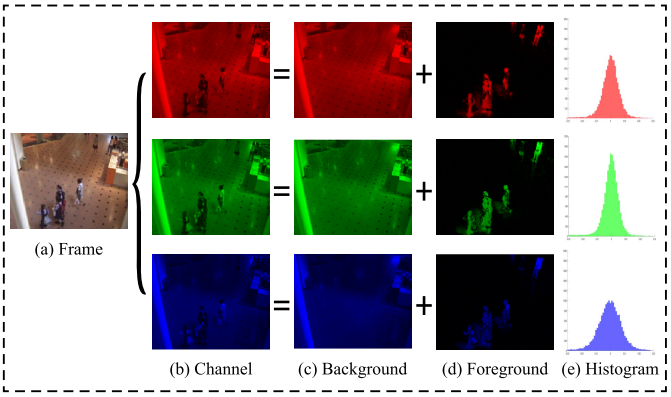


Fig. 8. Typical background subtraction result by our proposed NIID-MSL method and the corresponding histograms of residuals. (a) Typical RGB frame of the test surveillance video. (b) R, G, and B channels of the frame. (c)–(e) Backgrounds, foregrounds (residuals), and residual histograms of three channels at the same scale.

In this experiment, we utilize the F-measure as the quantitative metric for performance evaluation, which is calculated as follows:

$$\text{F-measure} = 2 \times \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

where precision $= (|\boldsymbol{S}_f \cap \boldsymbol{S}_{gt}|/|\boldsymbol{S}_f|)$ and recall $= (|\boldsymbol{S}_f \cap \boldsymbol{S}_{gt}|/|\boldsymbol{S}_{gt}|)$, $\boldsymbol{S}_f$ and $\boldsymbol{S}_{gt}$ denoted the support sets of the foreground calculated from the competing methods and the groundtruth one, respectively. For all competing methods, we choose many different threshold values to make the F-measure as large as possible. The larger the F-measure, the more accurate the foreground object is detected by the corresponding method.

The videos employed in this experiment contain different scenes, which involve not only simple static (such as Bootstrapping, Apertu, Escalator, and Lobby) but also complex dynamic (such as Camouflage, SwitchLight, WavingTrees, and WaterSurface) backgrounds. For videos with static backgrounds, we simply set the rank as 1 for all competing methods. However, for videos with dynamic backgrounds, we set the rank ranging from 1 to 5 and then determined the optimal rank parameter according to the F-Measures of all competing methods. It should be noted that, in most cases, all of the utilized methods attain the best performance under the same rank setting, and their performances are, thus, fairly tested under the same optimal rank parameter. More specifically, we set the rank parameter of all competing methods as three for videos of Camouflage, SwitchLight, TimeOfDay in Table IV and videos of Campus, Fountain, Hall in Table V, and set the rank as 5 for the other videos. Note that we regard the absolute value of the residual term $|\boldsymbol{X}^v - \boldsymbol{L}^v R|$ as the foreground to calculate the F-Measure.

Tables IV and V list the F-measure value calculated from all the competing methods of the two groups of data sets, respectively. It is easy to see that out proposed NIID-MSL achieves the best performance on average. More concretely,

[4]The true noise distribution curve is depicted using kernel density estimator, and the estimated distributions obtained by the NIID-MSL method are also plotted using kernel density estimator, but according to the estimated residuals.

[5]https://www.microsoft.com/en-us/research/project/test-images-for-wallflower-paper/

[6]http://vis-www.cs.umass.edu/ narayana/castanza/I2Rdataset/

TABLE IV

F-MEASURE VALUE OF ALL COMPETING METHODS ON THE WALLFLOWER DATA SET

Video	Methods					
	FLSSS	MSL	JIVE	CSRL	MISL	NIID-MSL
Bootstrapping	0.70	0.72	0.73	0.73	0.73	0.73
Camouflage	0.71	0.71	*0.73*	0.71	0.71	**0.74**
Apertu	0.85	**0.96**	0.57	0.86	0.86	**0.96**
SwitchLight	0.53	*0.67*	0.48	0.59	0.60	**0.68**
TimeOfDay	0.46	**0.76**	0.51	0.66	0.65	**0.76**
WavingTrees	*0.64*	0.61	0.62	0.61	0.63	**0.91**
Mean	0.65	*0.74*	0.61	0.69	0.70	**0.80**

TABLE V

F-MEASURE VALUE OF ALL COMPETING METHODS ON THE I2R DATA SET

Video	Methods					
	FLSSS	MSL	JIVE	CSRL	MISL	NIID-MSL
Campus	0.64	**0.73**	0.54	0.63	0.63	*0.72*
Escalator	0.66	0.68	0.68	0.69	*0.70*	**0.75**
Fountain	0.57	*0.82*	0.48	0.58	0.56	**0.83**
Curtain	0.73	*0.88*	0.46	0.83	0.83	**0.90**
Hall	0.60	*0.71*	0.57	0.60	0.62	**0.75**
Lobby	**0.93**	0.91	0.88	0.91	0.91	*0.92*
ShoppingMall	**0.85**	0.68	0.81	*0.84*	*0.84*	*0.84*
WaterSurface	0.81	*0.82*	0.80	*0.82*	*0.82*	**0.84**
Mean	0.72	*0.78*	0.65	0.74	0.74	**0.82**

NIID-MSL obtains a comparable or better result for the videos with a simple static background, such as Bootstrapping, Apertu, Escalator, ShoppingMall, and so on. This is because the latent subspace prior, which is strong enough to extract the background, equipped with a relatively robust loss (such as the L_1 loss in MSL, Cauchy loss in MISL) naturally leads to a good performance. However, for videos with complex dynamic background, such as waving tree in the WavingTrees and flickering computer screen in the Camouflage, the superiority of the NIID-MSL is more prominent.

Furthermore, we take the Wallflower data set to compare the running times of all competing methods. The results are listed in Table VI. The evaluation was performed in MATLAB R2018a on a computer with a 12-core Intel(R) Core i7-8700K CPU at 3.7 GHz and 16-GB RAM. From the table, it can be seen that compared with the other methods, the computational cost of the proposed method is at a moderate level overall. The FLSSS, JIVE, and CSRL method run faster than the MSL, MISL, and NIID-MSL methods, since they are constructed based on the L_2 loss, which is easy to handle during optimization. The MSL and MISL methods both introduce robust loss into the MSL: the former adopts L_1 loss and the latter uses Cauchy loss. Our proposed NIID-MSL, which is also focused on the robust issue, has a similar running speed compared with MSL and is evidently faster than MISL. Considering its better robust performance in all experiments, it should be reasonable to say that our proposed method is relatively efficient.

D. Background Subtraction on Multiview Data

In this section, our method is applied to the problem of background subtraction on videos captured from multiview cameras. Two multiview video sequences including one outdoor scene[7] (Fig. 9) and one indoor scene[8] (Fig. 10) are employed, which are captured by three cameras located at different positions targeting at the similar scene. Each surveillance video contains thousands of frames. Due to the memory limitation of our computer, we only extracted about 200 frames (taking one from every five frames) to compose our evaluation data.

TABLE VI

RUNNING TIMES (SECONDS) OF ALL COMPETING METHODS ON WALLFLOWER DATA SET

Video	Methods					
	FLSSS	MSL	JIVE	CSRL	MISL	NIID-MSL
Bootstrapping	1.92	30.56	13.80	2.07	96.08	12.62
Camouflage	8.24	35.41	11.00	1.93	280.57	44.41
Apertu	1.91	30.49	5.62	2.02	98.31	42.61
SwitchLight	5.40	23.60	9.65	1.24	194.35	31.71
TimeOfDay	9.62	40.34	14.00	2.11	326.46	47.08
WavingTrees	9.35	35.78	14.92	2.01	308.98	48.02
Mean	6.07	32.70	11.49	1.90	217.46	37.74

TABLE VII

F-MEASURE VALUE COMPARISON OF ALL COMPETING METHODS ON PETS09 DATA SET

Video	Methods					
	FLSSS	MSL	JIVE	CSRL	MISL	NIID-MSL
View1	0.64	*0.74*	0.63	0.67	0.65	**0.75**
View2	0.65	*0.87*	0.67	0.75	0.71	**0.88**
View3	0.71	**0.87**	0.64	0.70	0.65	**0.87**

TABLE VIII

F-MEASURE VALUE COMPARISON OF ALL COMPETING METHODS ON LAB DATA SET

Video	Methods					
	FLSSS	MSL	JIVE	CSRL	MISL	NIID-MSL
View1	0.75	*0.82*	0.80	0.80	0.80	**0.84**
View2	0.73	0.77	0.65	0.77	0.77	**0.80**
View3	0.53	**0.82**	0.64	0.56	0.57	*0.79*

The FLSSS, MSL, CSRL, JIVE, MISL, and our proposed NIID-MSL were then implemented on the data set for the task. We first ran each method on multiview videos to get the low-rank components as the background. Then, we obtained the foreground by calculating the absolute values of differences between the original frames and the estimated backgrounds. The rank was set as 3 for all competing methods for a fair comparison. The small rank setting is due to the fact that the tested videos are captured from statistic scenes, and their backgrounds are almost fixed and with a strong correlation.

The results of the subtracted backgrounds and foregrounds on several typical frames by all competing methods are shown in Figs. 9 and 10 for easy comparison. It can be observed that all the competing methods can extract the background from the videos with a slight difference in visualization, while the background images achieved by our method preserve relatively more elaborate details. In particular, for the outdoor scene video (Fig. 9), in which some people walked around the pic-

[7] http://cs.binghamton.edu/ mrldata/pets2009

[8] https://cvlab.epfl.ch/data/data-pom-index-php/

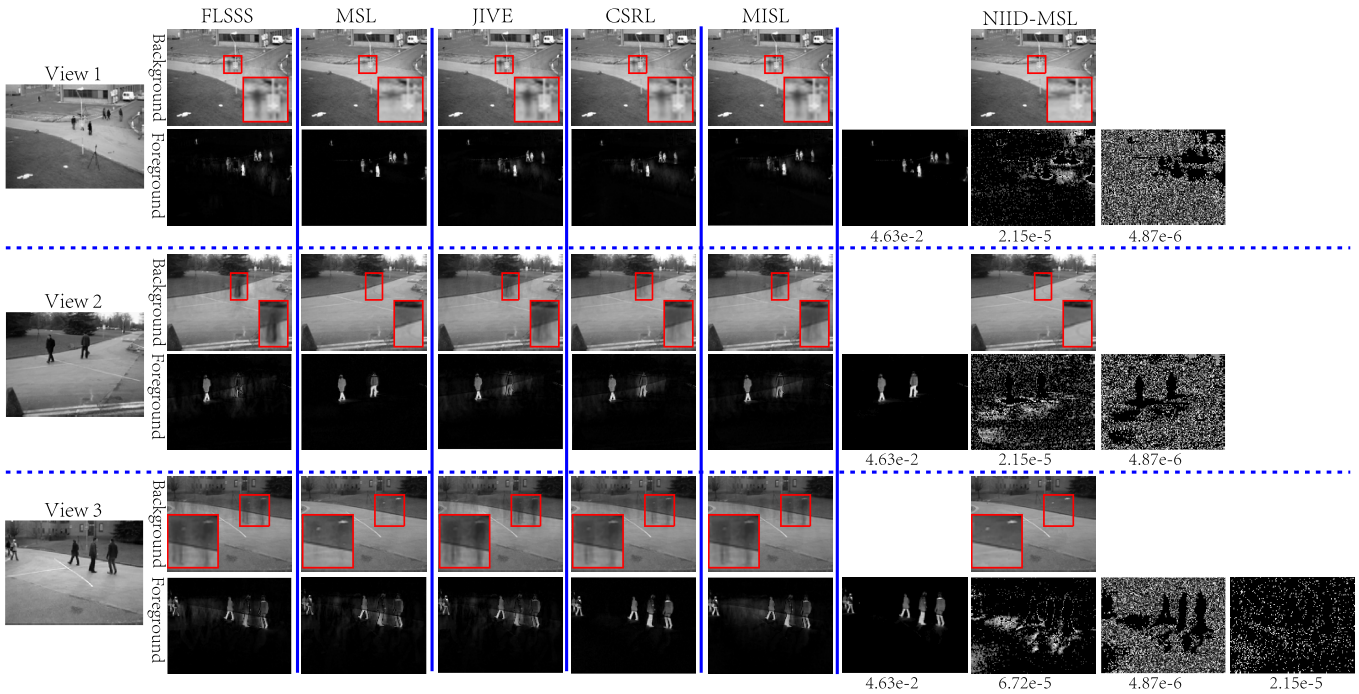


Fig. 9. Original video frames of three views of the Pets09 data set, background extracted by FLSSS, MSL, JIVE, CSRL, MISL, and NIID-MSL, together with their extracted noise images (left to right). For NIID-MSL, there are four Gaussian components to fit the noise at the top-level of HDP, and 3, 3, and 4 of which appear in the second level to noise of View 1, View 2, and View 3, respectively. The number below the foreground subtracted by the NIID-MSL method is the variance of the Gaussian components in the corresponding views.

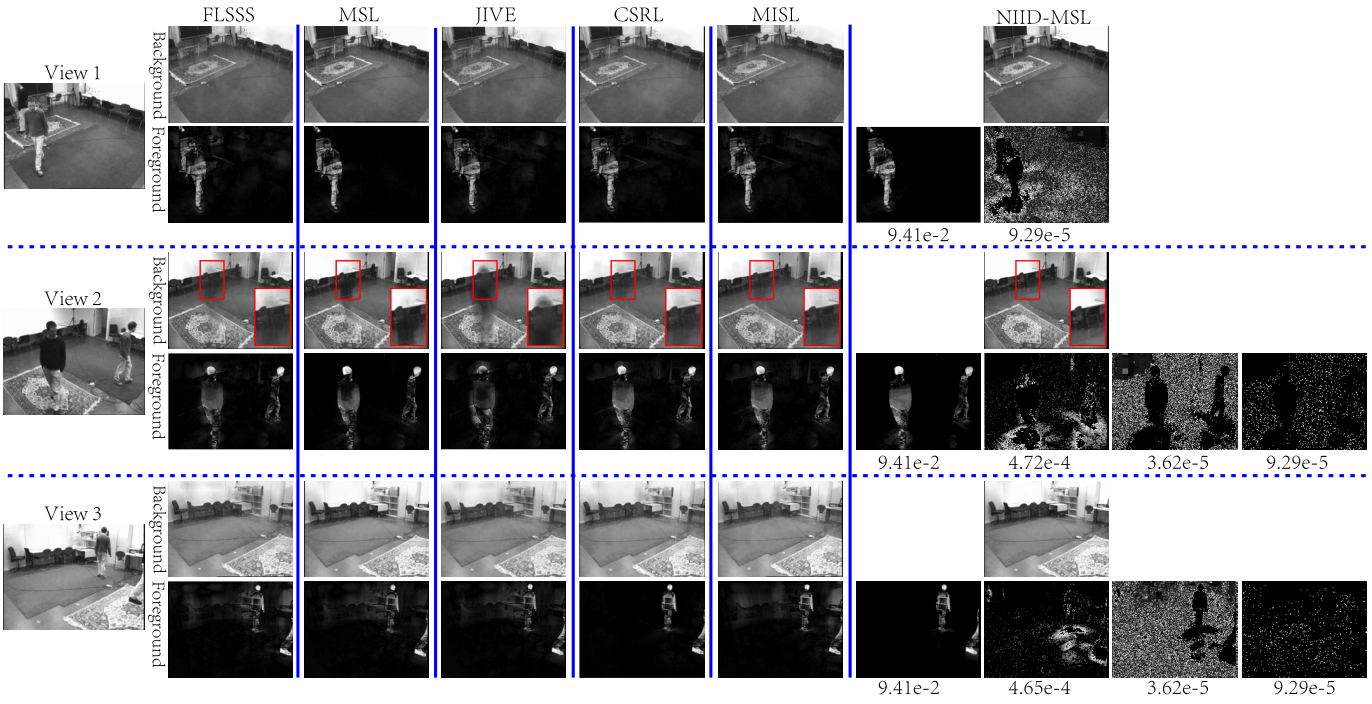


Fig. 10. Original video frames of three views of the Lab data set, background extracted by FLSSS, MSL, JIVE, CSRL, MISL, and NIID-MSL, together with their extracted noise images (left to right). For NIID-MSL, there are five Gaussian components at the top-level of HDP, and 2, 4 and 4 of which appear in the second level to characterize the noise of View 1, View 2, and View 3, respectively. The number below the foreground subtracted by NIID-MSL method is the variance of the Gaussian components in the corresponding views.

tured scene, our proposed NIID-MSL method can effectively separate them absolutely as foreground so as to get perfect background. As a comparison, other competing methods all extract the background in a relatively coarse manner with varying degrees of human shadow, since their simple i.i.d. noise assumption does not finely fit the noise situations in

TABLE IX

VARIANCES OF THE GAUSSIAN COMPONENTS EXTRACTED BY OUR METHOD IN EACH VIEW OF THE OUTDOOR SCENE VIDEO

Views	Variance 1	Variance 2	Variance 3	Variance 4
View 1	4.63e-2	2.15e-5	4.87e-6	
View 2	4.63e-2	2.15e-5	4.87e-6	
View 3	4.63e-2	6.72e-5	4.87e-6	2.15e-5

TABLE X

VARIANCES OF THE GAUSSIAN COMPONENTS EXTRACTED BY OUR METHOD IN EACH VIEW OF THE INDOOR SCENE VIDEO.

Views	Variance 1	Variance 2	Variance 3	Variance 4
View 1	9.41e-2	9.29e-5		
View 2	9.41e-2	4.72e-4	3.62e-5	9.29e-5
View 3	9.41e-2	4.65e-4	3.62e-5	9.29e-5

all three practical scenarios. Furthermore, in order to quantify the comparison result, we carefully annotated the foreground mask of one typical frame in each data set and calculated the F-Measure of each view. The detailed F-Measure values are shown in Tables VII and VIII, and the visualization corresponds to Figs. 9 and 10, respectively. It is easy to see the superiority of our proposed method, both in quantitative metrics and visualization.

A more interesting observation is that our NIID-MSL method is able to discover several modalities of the foreground information, through using GMM to fit the residuals in each view. The components of GMM with different variances reflect the noises with different scales and extents. What is more, the GMM noise extracted from each view share some components of the top-level GMM of HDP by our method. Thus, the variances of the Gaussian components in different views may or may not be the same, which depicts the correlations and distinctiveness of the noise in different views. Here, we list the variance of the Gaussian components of each view in these two groups of experiments in Tables IX and X, respectively. In addition, each component of GMM noise is visualized in Figs. 9 and 10 and their corresponding variance is also marked. These results apparently illustrate and substantiate the better noise fitting capability of our proposed method.

VII. CONCLUSION

We have proposed a new MSL method by sharing noise among different views under the Bayesian framework. Compared with most of the current MSL methods, which assume simple i.i.d. noise distribution (e.g., Gaussian or Laplacian) on all views of data, our method simultaneously considers the distinctiveness and correlation of noises among different views of data and models the noise of each view as a DPGMM with its specific parameters. It can perform the MSL task with better noise fitting capability in practical applications and shows more robust performance in the presence of complex noise. The effectiveness of our method is substantiated by experiments implemented on synthetic and real data sets. This paper paves the path for modeling complex noises in machine learning, and the generalization of such noise modeling idea to other practical problems such as multiview text clustering/classification should be a meaningful research direction for future research.

REFERENCES

[1] B. Wu, H. Ai, C. Huang, and S. Lao, "Fast rotation invariant multi-view face detection based on real Adaboost," in *Proc. Proc. 6th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2004, pp. 79–84.

[2] A. Mittal and L. S. Davis, "M_2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene," *Int. J. Comput. Vis.*, vol. 51, no. 3, pp. 189–203, 2003.

[3] H. Yong, D. Meng, W. Zuo, and L. Zhang, "Robust online matrix factorization for dynamic background subtraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 7, pp. 1726–1740, Jul. 2018.

[4] V. H. Hiep, R. Keriven, P. Labatut, and J.-P. Pons, "Towards high-resolution large-scale multi-view stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1430–1437.

[5] C. Wang, Y. Wang, Z. Lin, A. L. Yuille, and W. Gao, "Robust estimation of 3D human poses from a single image," in *Proc. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2361–2368.

[6] Z. Lin, C. Xu, and H. Zha, "Robust matrix factorization by majorization minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 208–220, Jan. 2017.

[7] C. Xu, D. Tao, and C. Xu, "Multi-view intact space learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 12, pp. 2531–2544, Dec. 2015.

[8] S. Sun, "A survey of multi-view machine learning," *Neural Comput. Appl.*, vol. 23, nos. 7–8, pp. 2031–2038, 2013.

[9] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view learning overview: Recent progress and new challenges," *Inf. Fusion*, vol. 38, pp. 43–54, Nov. 2017.

[10] Y. Jia, M. Salzmann, and T. Darrell, "Factorized latent spaces with structured sparsity," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 982–990.

[11] M. White, X. Zhang, D. Schuurmans, and Y.-L. Yu, "Convex multi-view subspace learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1673–1681.

[12] Y. Zhao, X. You, Y. Wei, S. Yin, D. Tao, and Y.-M. Cheung, "Multi-view latent space learning based on local discriminant embedding," in *Proc. 7th Int. Conf. Cloud Comput. Big Data (CCBD)*, Nov. 2016, pp. 225–230.

[13] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1888–1902, Aug. 2018.

[14] Y. Guo, "Convex subspace representation learning from multi-view data," in *Proc. 27th AAAI Conf. Artif. Intell.*, 2013, pp. 387–393.

[15] X. Cao, Q. Zhao, D. Meng, Y. Chen, and Z. Xu, "Robust low-rank matrix factorization under general mixture noise distributions," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4677–4690, Oct. 2016.

[16] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and Y. Yan, "L_1-norm low-rank matrix factorization by variational Bayesian method," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 4, pp. 825–839, Apr. 2015.

[17] R. Wang, B. Chen, D. Meng, and L. Wang, "Weakly-supervised lesion detection from fundus images," *IEEE Trans. Med. Imag.*, to be published.

[18] Z. Yue, D. Meng, Y. Sun, and Q. Zhao, "Hyperspectral image restoration under complex multi-band noises," *Remote Sens.*, vol. 10, no. 10, p. 1631, 2018.

[19] D. Meng and F. De La Torre, "Robust matrix factorization with unknown noise," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 1337–1344.

[20] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang, "Robust principal component analysis with complex noise," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 55–63.

[21] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, nos. 3–4, pp. 321–377, 1936.

[22] S. M. Kakade and D. P. Foster, "Multi-view regression via canonical correlation analysis," in *Proc. Int. Conf. Comput. Learn. Theory*. Berlin, Germany: Springer, 2007, pp. 82–96.

[23] F. R. Bach and M. I. Jordan, "A probabilistic interpretation of canonical correlation analysis," Univ. California, Berkeley, CA, USA, Tech. Rep. 688, 2005.

[24] S. Akaho. (2006). "A kernel method for canonical correlation analysis." [Online]. Available: https://arxiv.org/abs/cs/0609071

[25] C. Archambeau and F. R. Bach, "Sparse probabilistic projections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 73–80.

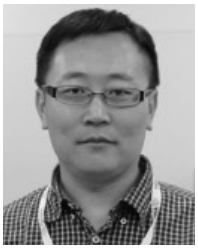
[26] M. A. Nicolaou, Y. Panagakis, S. Zafeiriou, and M. Pantic, "Robust canonical correlation analysis: Audio-visual fusion for learning continuous interest," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 1522–1526.
[27] C. H. Ek, P. H. S. Torr, and N. D. Lawrence, "Gaussian process latent variable models for human pose estimation," in *Proc. Int. Workshop Mach. Learn. Multimodal Interact.* Berlin, Germany: Springer, 2007, pp. 132–143.
[28] A. Shon, K. Grochow, A. Hertzmann, and R. P. Rao, "Learning shared latent structure for image synthesis and robotic imitation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 1233–1240.
[29] R. S. Cabral, F. Torre, J. P. Costeira, and A. Bernardino, "Matrix completion for multi-label image classification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 190–198.
[30] C. Christoudias, R. Urtasun, and T. Darrell. (2012). "Multi-view learning in the presence of view disagreement." [Online]. Available: https://arxiv.org/abs/1206.3242
[31] A. Goldberg, B. Recht, J. Xu, R. Nowak, and X. Zhu, "Transduction with matrix completion: Three birds with one stone," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 757–765.
[32] I. Mizera and C. H. Müller, "Breakdown points of cauchy regression-scale estimators," *Statist. Probab. Lett.*, vol. 57, no. 1, pp. 79–89, 2002.
[33] E. F. Lock, K. A. Hoadley, J. S. Marron, and A. B. Nobel, "Joint and individual variation explained (JIVE) for integrated analysis of multiple data types," *Ann. Appl. Statist.*, vol. 7, no. 1, pp. 523–542, 2013.
[34] G. Zhou, A. Cichocki, Y. Zhang, and D. P. Mandic, "Group component analysis for multiblock data: Common and individual feature extraction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2426–2439, Nov. 2016.
[35] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *Ann. Statist.*, vol. 1, no. 2, pp. 209–230, 1973.
[36] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statist. Sinica*, vol. 4, no. 2, pp. 639–650, 1994.
[37] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1566–1581, Dec. 2006.
[38] C. E. Rasmussen, "The infinite Gaussian mixture model," in *Proc. Adv. Neural Inf. Process. Syst.*, 2000, pp. 554–560.
[39] P. Chen, N. Wang, N. L. Zhang, and D.-Y. Yeung, "Bayesian adaptive matrix factorization with automatic model selection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1284–1292.
[40] M. West, "Hyperparameter estimation in Dirichlet process mixture models," Duke Univ., Durham, NC, USA, ISDS Discussion Paper #92-A03, 1992.
[41] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse Bayesian methods for low-rank matrix estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3964–3977, Aug. 2012.
[42] R. M. Neal, "Markov chain sampling methods for Dirichlet process mixture models," *J. Comput. Graph. Statist.*, vol. 9, no. 2, pp. 249–265, Jun. 2000.
[43] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer, 2006.
[44] D. M. Blei and M. I. Jordan, "Variational inference for Dirichlet process mixtures," *Bayesian Anal.*, vol. 1, no. 1, pp. 121–143, 2006.
[45] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
[46] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, 2010.
[47] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 1, Sep. 1999, pp. 255–261.

Zongsheng Yue received the B.Sc. degree in applied mathematics from Xinjiang University, Ürümqi, China, in 2015. He is currently pursuing the Ph.D. degree with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China, under the supervision of Prof. D. Meng.

His current research interests include noise modeling and image restoration.

Hongwei Yong received the B.Sc. and M.Sc. degrees from Xi'an Jiaotong University, Xi'an, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong.

His current research interests include image modeling and deep learning.



Deyu Meng (M'13) received the B.Sc., M.Sc., and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 2001, 2004, and 2008, respectively.

From 2012 to 2014, he was a Visiting Scholar with Carnegie Mellon University, Pittsburgh, PA, USA. He is currently a Professor with the Institute for Information and System Sciences, Xi'an Jiaotong University. His current research interests include self-paced learning, noise modeling, and tensor sparsity.

Qian Zhao received the B.Sc. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 2009 and 2015, respectively.

From 2014 to 2015, he was a Visiting Scholar with Carnegie Mellon University, Pittsburgh, PA, USA. He is currently a Lecturer with the School of Mathematics and Statistics, Xi'an Jiaotong University. His current research interests include low-rank matrix/tensor analysis, Bayesian modeling, and self-paced learning.

Yee Leung received the B.S.Sc. degree in geography from The Chinese University of Hong Kong, Hong Kong, in 1972, the M.S. and Ph.D. degrees in geography, and the M.S. degree in engineering from the University of Colorado Boulder, Boulder, CO, USA, in 1974, 1977, and 1977, respectively.

He is currently the Emeritus Professor with the Department of Geography and Resource Management and an Honorary Senior Research Fellow with the Institute of Future Cities, The Chinese University of Hong Kong. His current research interests include the development and applications of intelligent spatial decision support systems, spatiotemporal data mining, fuzzy sets and logic, rough set, neural networks, and environmental change.

Dr. Leung serves in the editorial boards of several international journals and chaired commissions in a number of professional associations.

Lei Zhang (M'04–SM'14) received the Ph.D. degree from Northwestern Polytechnical University, Xi'an, China, in 2001.

From 2001 to 2002, he was a Research Associate with the Department of Computing, The Hong Kong Polytechnic University. From 2003 to 2006, he was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. In 2006, he joined the Department of Computing, The Hong Kong Polytechnic University, as an Assistant Professor, where he has been the Chair Professor since 2017. His current research interests include computer vision, pattern recognition, and biometrics.

Dr. Zhang served as an Associate Editor of *SIAM Journal of Imaging Sciences*, the IEEE TRANSACTION ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and *Image and Vision Computing*. He is currently a Senior Associate Editor of the IEEE TRANSACTION ON IMAGE PROCESSING.